Linear Servo Theory

By ROBERT E. GRAHAM

The servo system is a special type of feedback amplifier, usually including a mixture of electrical, mechanical, thermal, or hydraulic circuits. With suitable design, the behavior of these various circuits can be described in the universal language of linear systems. Further, if the servo system is treated in terms of circuit response to sinusoidally varying signals, it then becomes possible to draw upon the wealth of linear feedback amplifier design based on frequency analysis.

This paper discusses a typical analogy between electrical and mechanical systems and describes, in frequency response language, the behavior of such common servo components as motors, synchro circuits, potentiometers, and tachometers. The elementary concepts of frequency analysis are reviewed briefly, and the familiar Nyquist stability criterion is applied to a typical motor-drive servo system. The factors to be considered in choosing stability margins are listed—system variability, noise enhancement, and transient response. The basic gain-phase interrelations shown by Bode are summarized, and some of their design implications discussed. In addition to the classical methods, simple approximate methods for calculating dynamic response of servo systems are presented and illustrated.

Noise in the input signal is discussed as a compelling factor in the choice of servo loop characteristics. The need for tailoring the servo loop to match the input signal is pointed out, and a performance comparison given for two simple servos designed to track an airplane over a straight line course. The use of subsidiary or local feedback to linearize motor-drive systems, and predistortion of the input signal to reduce overall dynamic error are described.

1. Introduction

A SIMPLE servo system is one which controls an output quantity according to some required function of an input quantity. This control is of the "report back" type. That is, some property of the output is monitored and compared against the input quantity, producing a net input or "error" signal which is then amplified to form the output. The first statement defines the servo as a transmission system; the second, as a feedback loop. The problem of servo design is then to fashion the desired transmission properties while meeting the stability requirements of the feedback loop. This is the familiar design problem of the feedback amplifier.

2. The Servo Circuit

The design of linear feedback amplifiers has been developed to a high degree in terms of frequency response; that is, in terms of circuit response to sinusoidal signals.[1] The servo system is a special type of feedback amplifier, and usually can be made fairly linear. Thus, it is logical to analyze and design the servo circuit on a frequency response basis. Also,

[1] See "Network Analysis and Feedback Amplifier Design," by H. W. Bode, D. Van Nostrand Co., 1945.

servo systems usually are combinations of electrical, mechanical, thermal, or hydraulic circuits. In order to describe the behavior of these various circuits in homogeneous terms, it is desirable to recognize the analogous relationships established by similarity of the underlying differential equations. Before proceeding to a discussion of frequency analysis, a typical analogy between electrical and mechanical systems will be described.

2.1 *Electrical-Mechanical Analogy*

Confining the discussion to rotating mechanical systems, the analogy which will be chosen here puts voltage equivalent to torque, and current to rotational speed. This choice leads to the array of equivalents shown in Fig. 1; inductance, capacity, and resistance corresponding to inertia, compliance, and mechanical resistance, respectively. Charge is equivalent to angular displacement, and both kinetic and potential energy are self-analogous. The ratio of voltage to current, or torque to speed has the dimensions of resistance. In an interconnected electro-mechanical system, the ratio of voltage to speed or torque to current may be called a transfer resistance. Similarly, the ratio of voltage to angular displacement, or of torque to charge, is a transfer stiffness (reciprocal of capacity or compliance).

Some commonly used devices for coupling between electrical and mechanical circuits are shown in Figs. 2 and 3. The motor, Fig. 2a, is used to convert an electrical current i into a mechanical speed or "current" $\dot{\theta}$ ($= d\theta/dt$). The electrical control current i is produced by the voltage difference between an applied emf e and a counter-rotational emf (not indicated), acting upon the total electrical mesh resistance R_e.* In the mechanical circuit, a torque proportional to i forces a "current" $\dot{\theta}$ through the mechanical load R_m, J.

An equivalent mechanical mesh directly relating shaft speed to the applied emf is shown in Fig. 2b. A fictitious generated torque $\mu_t e$ acts upon the mechanical load through an apparent mechanical resistance R_m'. μ_t is a transfer constant determined both by the motor properties and the electrical mesh resistance R_e. R_m' is similarly governed and is inversely proportional to R_e.

The motor may be compared to a vacuum tube having an amplification factor μ_t and a plate resistance R_m'. However, the motor is usually much more a bilateral coupling element than the vacuum tube, due to the effect of the counter emf upon the electrical mesh.

The potentiometer, tachometer, and synchro circuit shown in Fig. 3 are all means for converting a mechanical quantity to an electrical one. All three are substantially unilateral coupling elements. The potentiometer

* R_e includes both the source resistance and the motor winding resistance.

delivers an output voltage e proportional to its shaft angle θ. Thus, the ratio of e to θ is a transfer stiffness constant S_t. The synchro circuit consists of a synchro generator connected to a control transformer, and delivers an output voltage e proportional to $\theta_1 - \theta_2$, the angular difference between

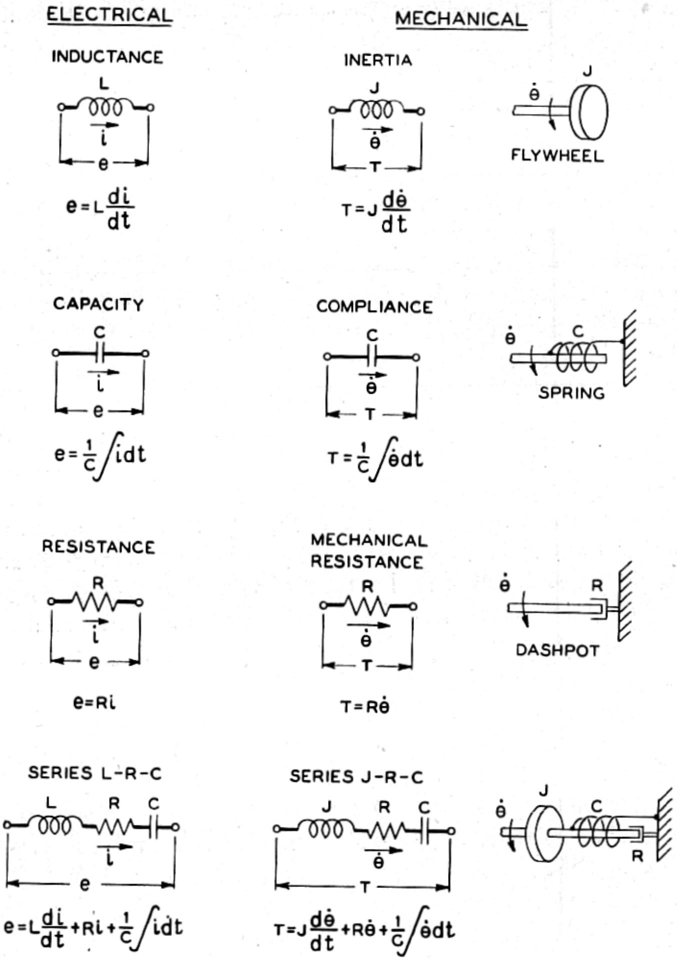

Fig. 1—Electrical-mechanical analogy.

the two shafts. Thus the action of the synchro pair is that of a combined transfer stiffness and differential. The tachometer is a generator which produces an output voltage e proportional to $\dot{\theta}$, the angular speed of its shaft. The ratio of e to $\dot{\theta}$ is a transfer resistance constant R_t, There are many other specific devices used to convert from mechanical to electrical quantities. Most are equivalent to the potentiometer or synchro circuit, one such being a lobing radar antenna, which delivers a voltage proportional to an angular difference. A different, less widely used device is the accelerometer, a generator which delivers an output voltage proportional to the angular acceleration of its shaft. Its characteristic is that of a transfer inductance or inertia.

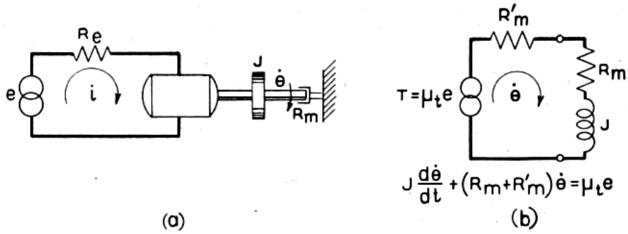

Fig. 2—Motor as a transfer device. (a) Motor and load. (b) Equivalent mechanical mesh.

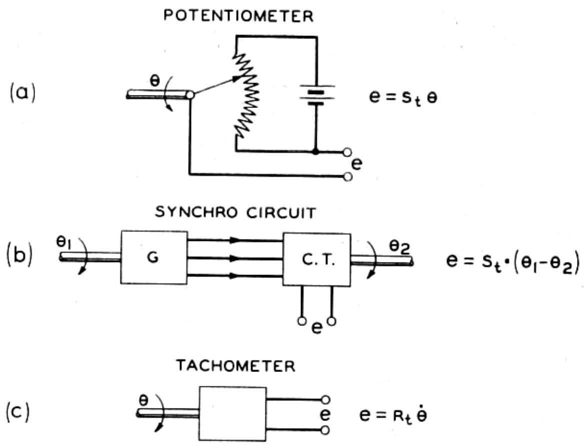

Fig. 3—Mechanical-electrical transfer devices.

2.2 *Frequency Analysis*

A brief review of the basic concepts of periodic analysis will be presented. It is assumed that the driving force applied to a network may be analyzed in terms of a series of sinusoidal components of various amplitudes, frequencies, and phases. The network response to each sinusoidal component is then evaluated, and the over-all result obtained by a summation of all such elementary responses. This is the formal procedure. Actually it is often unnecessary to perform these precise operations in order to obtain a broad picture of the network behavior.

The method for determining the network response to a sinusoidal input is developed as follows. It is assumed that the circuit parameters are constant, independent of signal amplitude. Then, as indicated in Fig. 1, a single R-L-C or R-J-C mesh may be represented by a constant-coefficient linear differential equation. Choosing the electrical mesh for illustration,

$$(Lp + R + 1/Cp)i(t) = e(t),$$

where $p^n = d^n/dt^n$, $1/p = \int dt$, and $e(t)$, $i(t)$ are the mesh voltage and current respectively. This may be further abbreviated as

$$Z(p)i(t) = e(t), \tag{1}$$

where $Z(p) = Lp + R + 1/Cp$. For purposes of frequency analysis we are interested only in the forced or steady-state solution of (1), where $e(t)$ is a sinusoidal voltage $E \sin \omega t$. This steady-state solution is

$$i(t) = \frac{E}{|Z(j\omega)|} \sin(\omega t + \phi),$$

where $j\omega$ has replaced p in the function $Z(p)$, and the phase shift ϕ is the negative of the phase angle of the complex number $Z(j\omega)$.[2] This result is conventionally abbreviated as

$$I = \frac{E}{Z(j\omega)}, \tag{2}$$

where I is a complex number whose magnitude equals the peak amplitude of the current, and whose phase angle gives the associated phase shift. The function $Z(j\omega)$ is called the impedance of the mesh.

The relationship between torque, angular speed, and mechanical impedance is of course the same as expressed by (2). That is,

$$\dot{\theta} = \frac{T}{Z(j\omega)}, \tag{2.1}$$

where $\dot{\theta}$ is the complex peak amplitude of the sinusoidally varying speed, T is the peak amplitude of the applied sinusoidal torque, and $Z(j\omega)$ is the mechanical impedance obtained by substituting $j\omega$ for p in the operator $Z(p) = Jp + R + 1/Cp$. Since the angular displacement is the time integral

[2] ω is used to represent frequency in radians/sec, or 2π times frequency in cycles/sec.

of the speed, the expression for θ may be obtained by dividing both sides of equation (2.1) by p or, for the periodic case, by $j\omega$. Thus

$$\theta = \frac{T}{j\omega Z(j\omega)}. \tag{2.2}$$

The function $j\omega Z(j\omega)$ is the complex stiffness of the mechanical mesh. The phase shift of θ relative to $\dot{\theta}$ is -90 degrees, as seen from a comparison of (2.1) and (2.2).

For an electro-mechanical network consisting of a number of interconnected meshes, a set of simultaneous differential equations of the type of (1) may be written. If $j\omega$ is substituted for p in these equations, there results a set of simultaneous algebraic equations which lead directly to the steady-state periodic solution. If a sinusoidal voltage or torque is applied at some mesh of the network, the resulting sinusoidal current or speed response in some other mesh is given by, using the notation of (2),

$$(\text{Response}) = \frac{(\text{Cause})}{Z_t(j\omega)},$$

where $Z_t(j\omega)$ for the chosen pair of meshes is obtained from the solution of the algebraic equations. $Z_t(j\omega)$ is called a transfer impedance, and may express the ratio of a voltage to current or speed, or of a torque to current or speed. The above relation also may be written as

$$(\text{Response}) = Y_t(j\omega)\cdot(\text{Cause}),$$

where $Y_t(j\omega) = 1/Z_t(j\omega)$ is called the transfer admittance between the two chosen parts of the network. In this form the response amplitude is obtained by multiplying the forcing sinusoid by $|Y_t(j\omega)|$, while the phase shift is given directly by the angle of $Y_t(j\omega)$. The transfer ratio between like or analogous quantities at two parts of the network is similarly a complex function of frequency, having the dimensions of a pure numeric.

Servo systems usually consist largely of elementary networks isolated by unilateral coupling devices (vacuum tubes, potentiometers, etc.). Thus, over-all transfer ratios often may be evaluated by taking the product of a number of simple "stage" transfer ratios, rather than by solving a large array of simultaneous equations. If the absolute magnitudes of the transfer ratios or "transmissions" are expressed in decibels[3] of logarithmic gain, both the over-all gain and phase shift of a number of tandem stages may be obtained by simple addition of the individual stage gain and phase values.

The transfer ratios of the conversion devices shown in Figs. 2 and 3

[3] The gain in decibels for a given transfer ratio is taken to be 20 $\log_{10}$ of the absolute value of the ratio.

may be written by inspection. Referring to Fig. 2b, the transfer admittance of a motor with resistance and inertia load may be written as

$$\frac{\dot{\theta}}{E} = \frac{\mu_t}{j\omega J + R_m + R_m'},$$

$$= \frac{\mu_t}{J} \cdot \frac{1}{j\omega + \omega_m}, \tag{3}$$

where

$$\omega_m = \frac{R_m + R_m'}{J}.$$

ω_m is the reciprocal of the time-constant of the motor and control mesh, and is 2π times the "corner" frequency at which the inertial impedance just equals the apparent mechanical resistance. Writing the transfer characteristic in terms of shaft position, rather than speed,

$$\frac{\theta}{E} = \frac{\mu_t}{J} \cdot \frac{1}{j\omega(j\omega + \omega_m)}. \tag{3.1}$$

For values of ω small compared with ω_m, θ/E is proportional to $1/j\omega$. This factor has a phase shift of -90 degrees and approaches infinity as ω approaches zero. This is merely a statement in frequency analysis language that the motor shaft angle is the time integral of the applied voltage, for slowly changing voltage. For more rapidly varying voltage, such that ω is large compared to ω_m, θ/E is proportional to $1/(j\omega)^2$ or $-1/\omega^2$, the angular variation being shifted -180 degrees with respect to the voltage variation.

The transfer ratio of the potentiometer, Fig. 3a, may be written as

$$\frac{E}{\theta} = S_t, \tag{4.1}$$

while for the tachometer, Fig. 3c,

$$\frac{E}{\dot{\theta}} = R_t, \tag{4.2}$$

or

$$\frac{E}{\theta} = j\omega R_t. \tag{4.3}$$

Usually at some point in the system a compensating or "equalizing" network will be included to modify the transfer ratio of the basic components to the desired over-all transmission characteristic. Frequently this equalizer is incorporated in the electrical section of the servo because of the comparative ease with which electrical circuit components may be assembled in desired combinations. The transfer characteristic of the equalizer may be simple or complicated, but in general may be written in the form,

$$\mu_e \sim \frac{(j\omega + \omega_1)(j\omega + \omega_3) \cdots}{(j\omega + \omega_2)(j\omega + \omega_4) \cdots}, \tag{5}$$

where the constants ω_1, ω_2, etc. may be real or complex. The synthesis of equalizing networks is a well known art and will not be discussed here, particularly since most of the equalization characteristics used in present servo systems can be realized with simple networks.

2.3 *Simple Servo System (Single Loop)*

The simple servo system may be divided into two basic parts, an amplifying circuit and a monitoring or comparison circuit. Such a division is

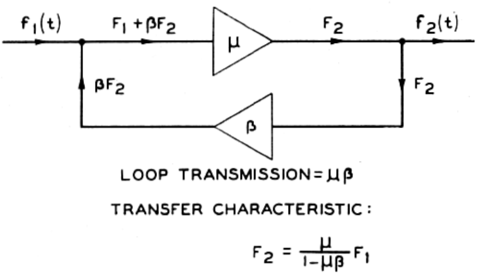

Fig. 4—Simple servo system.

indicated in Fig. 4, where μ and β are the transfer characteristics of the amplifying and monitoring parts, respectively. F_1 and F_2 represent typical sinusoidal components of the total input and output quantities $f_1(t)$ and $f_2(t)$,[4] while μ and β are complex-valued functions of $j\omega$ as described in the previous section.

The return signal βF_2 from the monitoring circuit is added to the servo input F_1 to form a net μ circuit input $F_1 + \beta F_2$. The servo transfer characteristic is found by setting

$$F_2 = \mu(F_1 + \beta F_2),$$

from which

$$F_2 = \frac{\mu}{1 - \mu\beta} F_1. \tag{6}$$

[4] That is, F_1 and F_2 are complex quantities employed in the same fashion as I in equation (2).

The closed system formed by the two basic circuits in tandem is of course a feedback loop, the loop transmission characteristic being given by $\mu\beta$.

Any desired form of servo transfer ratio may be obtained by an unlimited number of μ and β circuit combinations.[5] However, the β characteristic, which is usually determined by a passive network or an inherent property of a monitoring device, tends to be more stable with time and varying signal amplitude than that of the μ circuit, which may include vacuum tubes, motors, and other variable components. Consequently, it is desirable to have the servo transfer characteristic largely dependent upon the β circuit properties alone. This may be accomplished by making the loop transmission $\mu\beta$ very large compared to unity over the essential frequency range of the servo input signal $f_1(t)$. Under this condition, equation (6) becomes,

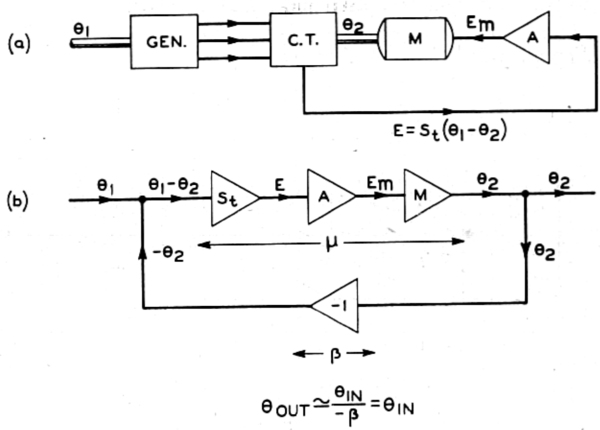

Fig. 5—Synchro follow-up system.

$$F_2 \simeq \frac{F_1}{-\beta}, \qquad |\mu\beta| \gg 1. \tag{6.1}$$

Thus the external transfer characteristic is set by β.* If, for instance, F_1 and F_2 are similar or analogous quantities and it is desired to have the servo output a replica of the input, β may be chosen as -1, yielding $F_2 \simeq F_1$.

It is not always easy to determine the basic parts μ and β of a servo by inspection of a schematic diagram of the system. An example is furnished by the synchro follow-up system shown in Fig. 5a. As previously discussed, the characteristic of the synchro comparison circuit is that of a differential

[5] Feedback stability requirements place certain restrictions on the permissible forms of $\mu\beta$. This will be discussed in the next section.

* The error arising from the approximate nature of (6.1) will be discussed in the next section as one type of "servo error."

transfer stiffness S_t, the voltage output of the control transformer being given by $S_t(\theta_1 - \theta_2)$. However, as seen from the modified diagram of Fig. 5b, the β characteristic is simply -1, the transfer constant S_t appearing in the μ circuit. Thus if the loop transmission is kept large, the essential relation demanded between θ_1 and θ_2 does not depend upon the value of S_t

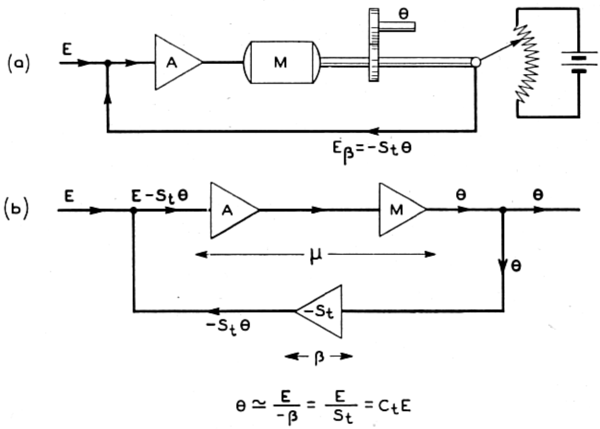

Fig. 6—Potentiometer loop.

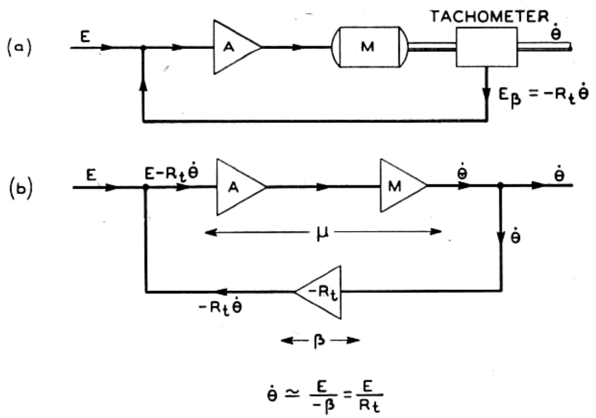

Fig. 7—Tachometer loop.

(as is obvious from physical considerations). This result also applies to a radar angle-tracking loop, where the received deviation or error signal is proportional to the difference between the angular coordinate of the target and that of the antenna system.

Figures 6 and 7 represent two servo systems where the input is electrical and the output mechanical. In Fig. 6 a potentiometer is used as a monitoring device, the transfer stiffness in this case appearing in the β circuit. If θ is regarded as the output, then $\beta = -S_t$,* and the transfer characteristic is, for high loop gain,

$$\theta \simeq \frac{E}{-\beta} = \frac{E}{S_t} = C_t E,$$

where $C_t = 1/S_t$. Thus the over-all characteristic between input voltage and angular displacement is simply a transfer compliance constant. In Fig. 7 a tachometer monitor is used. Regarding angular speed $\dot{\theta}$ as the output, then $\beta = -R_t$, the transfer resistance of the tachometer. The transfer equation is thus

$$\dot{\theta} \simeq \frac{E}{-\beta} = \frac{E}{R_t} = g_t E,$$

where $g_t (= 1/R_t)$ is a transfer conductance.

3. Design of Simple Linear Servo Systems

The majority of servo systems in use, while often greatly extended in space and frequently including highly diversified transmission elements, may be represented by one essential feedback loop. However, a well designed servo often will incorporate numerous subsidiary or local feedback loops around stages of the system, in order to obtain a desired degree of linearity or performance with easily obtainable circuit components. Common examples of such local feedback loops are electrical feedback around vacuum tube amplifiers, and tachometer (velocity) feedback around motor-drive systems. These subsidiary feedback loops are almost always designed so that they are individually stable when the over-all feedback loop is opened (assuming the method employed for opening the over-all loop does not disturb the impedance terminations of the local feedback stages). If it is thus assumed that any subsidiary loops are individually stable, then the primary servo loop design may be treated simply as that of a single loop, whose over-all loop transmission is obtained by taking the product of the external transfer ratios of the various stages.

The design of a single loop servo may be divided into the design of the loop transmission $\mu\beta$, and one of the remaining parts, μ or β. As previously described, it is usually desirable to fix β according to the required basic input-output relationship of the servo, thus leaving $\mu\beta$ as a single characteristic to be chosen.

* It is assumed here that the transmission of the μ circuit is basically positive. The negative signs associated with S_t of Fig. 6 and R_t of Fig. 7 are then introduced (by poling) to make the loop transmission $\mu\beta$ essentially negative. This stipulation ensures what is commonly called "negative feedback," when the loop delay is zero.

As usual, specification of the form of $\mu\beta(j\omega)$ is beset by a series of performance objectives on the one hand and a set of physical limitations and restrictions on the other. Assuming the relationship expressed by (6.1) to be the required one, it would seem desirable to make $\mu\beta(j\omega)$ very large compared to unity at all frequencies. However, there are reasons why this is neither possible nor actually desirable. As the value of ω is increased, the loop transmission is eventually controlled by parasitic circuit elements such as distributed capacity and inductance in the electrical circuits, and distributed inertia, compliance, and backlash in the mechanical circuits. The effect of these parasitic elements at the higher signal frequencies is to cause $|\mu\beta|$ to decrease as a very high order of $1/\omega$ with increasing frequency. It will be shown, however, that feedback stability considerations require the loop transmission to be decreasing comparatively slowly through the frequency region where $|\mu\beta|$ is of the order of unity. Thus $\mu\beta$ must be reduced below unity at a frequency sufficiently low to avoid excessive contribution from the parasitics.

The presence of "noise" or undesired disturbances in the servo input signal is another compelling factor in the design of the loop characteristic. Input noise is harmful both in causing spurious output fluctuations and in overloading the power stages of the servo system. Both of these effects are reduced by narrowing the frequency band of the servo transfer characteristic. Referring to the expression for the transfer characteristic given by (6), it may be seen that a restricted transfer bandwidth may be obtained by reducing μ and $\mu\beta$ well below unity at a small value of signal frequency.[6]

On the other side of the picture is the requirement of fidelity in maintaining the desired input-output relationship. Undue narrowing of the transfer bandwidth of the servo results in large dynamic error, the magnitude of which depends both upon the character of the input signal and upon the chosen transfer characteristic.

The optimum design of a servo system, for a specified input signal and noise, thus is a compromise between dynamic error and output noise fluctuations, with stability considerations and parasitic circuit elements restricting the possible choice of loop transmission characteristics.

3.1 *Stability of Single Loop Systems*

The word *stable* as applied to a servo system is used here to imply a system whose transient response decreases with increasing time. It is possible

[6] When the β characteristic is under suitable design control, another method is available. Thus if β is made to rise in the frequency region of the desired transfer cut-off, and if $\mu\beta$ is maintained large beyond this region, (6.1) shows that the desired restriction is effected. For a given transfer characteristic, this cut-off method requires a wider frequency range for $\mu\beta$ and is thus more vulnerable to the effects of parasitic circuit elements. However, shaping of both the μ and β circuits permits a more rapid cut-off of the servo transfer characteristic than is possible with μ circuit shaping alone.

to determine the stability of a completed servo design by obtaining the transient solution of its differential equation. Though often very tedious, this is fairly straightforward. However, this method of procedure often is of little help either in guiding the initial design or in predicting the necessary changes, should the trial design be found unstable. The addition of even one circuit element to a design will generally create an entirely new differential equation whose solution must be found.

An alternative method for determining servo system stability, based on frequency analysis, furnishes the necessary information in a form which greatly facilitates the design procedure. This method is relatively simple to apply, even when the system has a large number of meshes and a high order differential equation, and the additive effects of minor circuit modifications are easily evaluated.

The stability of a single loop servo system—or of a primary loop, when the subsidiary loops are individually stable—may be investigated by plotting the negative of the loop transmission, $-\mu\beta(j\omega)$, on a complex plane for real values of ω ranging from minus infinity to plus infinity. (The negative sign is introduced because the loop transmission $\mu\beta$ is generally arranged to have an implicit negative sign, apart from network phase shifts. Thus $-\mu\beta$ is a positive real number when the network phase shift is zero.) *Then the necessary and sufficient criterion for system stability is that the resulting closed curve must not encircle or intersect* the point $-1,0$.* This type of plot is commonly called a Nyquist diagram, and is widely used in the design of electrical feedback amplifiers. An added stipulation is necessary if $\mu\beta(j\omega)$ becomes infinite at a real value of ω, say ω'. In this case an infinitesimal positive real quantity ϵ must be added to $j\omega$; that is, the function to be plotted is $\mu\beta(j\omega + \epsilon)$. This has no effect upon the plot except in the neighborhood of the singularity, where $\mu\beta(j\omega + \epsilon)$ is caused to traverse an arc of infinite radius as ω is varied through the value ω'.

As seen from (3.1), inclusion of a motor in a servo loop of the type shown in Figs. 5 and 6 will cause an infinite loop transmission at $\omega = 0$, assuming there is transmission around the remainder of the loop at zero frequency. The motor is the only commonly encountered circuit element capable of producing an infinite loop transmission at real frequencies.

In order to illustrate the use of the Nyquist diagram, a motor servo system of the type shown in Fig. 6 will be chosen. Again referring to equation (3.1), it may be seen that the transfer ratio of the motor and potentiometer is,

* This criterion is due to H. Nyquist, "Regeneration Theory,". *B. S. T. J.*, January 1932. Detailed descriptions of stability criteria for single and multiple loop systems are given by Bode, loc. cit., and by L. A. MacColl, "Fundamental Theory of Servomechanisms," D. Van Nostrand Co., 1945.

$$\frac{E_\beta}{E_m} = \frac{-S_t \mu_t}{J} \cdot \frac{1}{j\omega(j\omega + \omega_m)}.$$

It is assumed that the amplifier includes an equalizing network such that the over-all amplifier characteristic is

$$A\left(\frac{j\omega + \omega_1}{\omega_1}\right)\left(\frac{\omega_2}{j\omega + \omega_2}\right)^3,$$

where A, ω_1, and ω_2 are positive real constants. The loop transmission is given by the product of these two transfer factors and thus may be written as

$$\mu\beta(j\omega) = -\frac{\omega_0}{j\omega}\left(\frac{\omega_m}{j\omega + \omega_m}\right)\left(\frac{j\omega + \omega_1}{\omega_1}\right)\left(\frac{\omega_2}{j\omega + \omega_2}\right)^3, \tag{7}$$

where ω_0 is a positive real constant given by

$$\omega_0 = \frac{AS_t \mu_t}{\omega_m J} = \frac{AS_t \mu_t}{R_m + R_m'}.$$

The three factors in parenthesis have been so grouped that they all approach unity for small values of ω. Thus the low-frequency behavior of $-\mu\beta$ is given by $\omega_0/j\omega$. This quantity has a pole at $\omega = 0$, so that it is necessary to plot the function

$$-\mu\beta(j\omega + \epsilon) \simeq \frac{\omega_0}{j\omega + \epsilon}, \tag{7.1}$$

in the neighborhood of $\omega = 0$. As ω is, in succession, a small negative quantity, zero, and a small positive quantity; the expression of (7.1) is correspondingly a large positive imaginary, a large positive real, and a large negative imaginary. Thus (7.1) traverses an infinite arc from the positive imaginary axis to the negative imaginary axis as ω increases through the value zero.

Assuming the numerical values

$$\begin{aligned} \omega_0 &= 200 \text{ sec}^{-1} \\ \omega_m &= 1 \quad `` \\ \omega_1 &= 10 \quad `` \\ \omega_2 &= 200 \quad `` , \end{aligned}$$

equation (7) may be rewritten as

$$-\mu\beta(j\omega) = \frac{200}{j\omega}\left(\frac{1}{j\omega + 1}\right)\left(\frac{j\omega + 10}{10}\right)\left(\frac{200}{j\omega + 200}\right)^3. \tag{7.2}$$

The phase angle of $-\mu\beta$ in degrees is, from (7.2),

$$B = -90 - \tan^{-1}\omega + \tan^{-1}\frac{\omega}{10} - 3\tan^{-1}\frac{\omega}{200}, \tag{7.3}$$

while the absolute magnitude is given by

$$|\mu\beta| = \frac{200}{\omega}\sqrt{\left(\frac{1}{1+\omega^2}\right)\left(\frac{10^2+\omega^2}{10^2}\right)\left(\frac{200^2}{200^2+\omega^2}\right)^3}. \tag{7.4}$$

The Nyquist diagram of (7.2) is shown in Fig. 8. To emphasize the important features, radial magnitudes have been plotted on a logarithmic

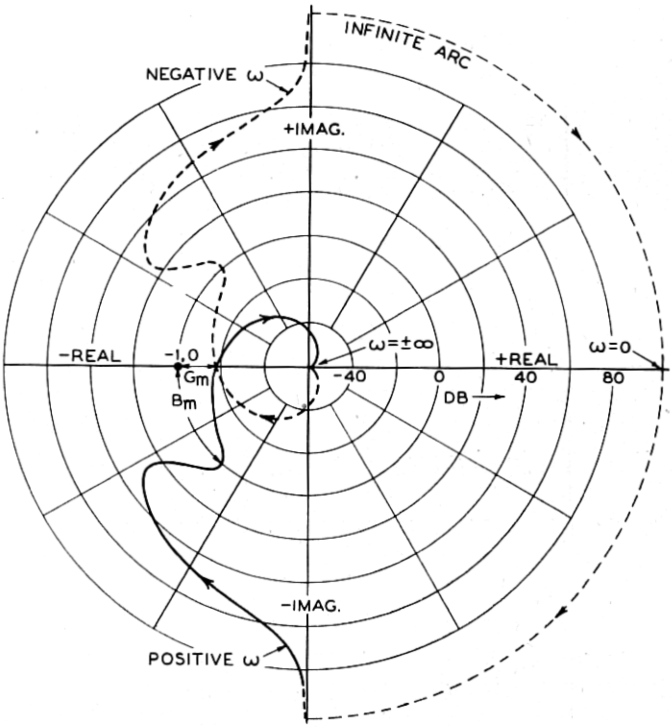

Fig. 8—Nyquist diagram of $-\mu\beta$.

scale.[7] The arrows indicate the direction of traversal as ω is varied from $-\infty$ to $+\infty$. The infinite arc traversed as ω varies through zero is indicated symbolically by the dotted semicircle in the right half plane.[8] As is the case for any physical system, the plot for negative values of ω is simply the mirror image of the positive frequency plot.

Since the polar plot does not encircle or intersect the "critical" point $-1,0$,

[7] Except in the immediate neighborhood of the origin, where a linear scale must be employed to plot the value $\mu\beta = 0$.

[8] The exact shape of this arc is of no consequence.

the system is seen to be stable.[9] From a practical standpoint it is necessary to know not only that a design is stable, but that it has sufficient margin against instability. The need for proper stability margin arises from two general considerations. First, the loop transmission of the physical system will vary with time due to aging, temperature changes, line voltage fluctuations, etc. Also the physical embodiment will depart from the paper design due to errors of adjustment and measurement, and to the effects of unallowed-for parasitic elements. Second, a design which is too near instability will have an undesirable transient response—large overshoots and persistent oscillations—and will unduly enhance noise in the input signal.

Stability margin is measured in a sense by the minimum displacement between the polar plot and the point $-1,0$. In feedback amplifier design, two numbers often are taken as a measure of margin against instability. These are called the ***gain margin*** and the ***phase margin***. The gain margin, G_m, measures the amount, in decibels, by which the magnitude of $\mu\beta$ falls short of unity, at a phase angle of ± 180 degrees. The numerical value of gain margin for the system of Fig. 8 is about 18 db, which is the required increase in amplifier gain to make the servo unstable.[10] That is, this increase in amplifier gain would multiply the curve of Fig. 8 by a constant factor such that it would intersect the point $-1,0$. The phase margin, B_m, is equal to the absolute magnitude of the angle between $-\mu\beta$ and the negative real axis, at $|\mu\beta| = 1$. Figure 8 illustrates a phase margin of about 50 degrees. That is, if the points on the curve where $|\mu\beta| = 1.0$ were swung toward the negative real axis by about 50 degrees they would coincide with the point $-1,0$, and the servo would be unstable.

The type of transient response obtained with reasonable gain and phase margins is indicated in Fig. 9, which shows the response of the illustrative servo system to an input step. The initial overshoot is about 25%, and the oscillation damps out very quickly. For the general case, (6) may be rewritten in the form

$$F_2 = \left[\frac{-\mu\beta}{1-\mu\beta}\right]\cdot\frac{F_1}{-\beta}. \tag{8}$$

The relation $F_2 = F_1/-\beta$ may be regarded as the desired one, with the bracketed factor acting as the inevitable modifier. Then if the quantity

[9] With more complicated systems it may not be obvious whether or not the plot encircles $-1, 0$. A simple test employs a vector with its origin at the $-1, 0$ point and its tip on the curve. If the vector undergoes zero net rotation as it traces along the curve from $\omega = 0$ to $\omega = \infty$, the curve does not encircle the critical point.

[10] In some servo systems a *decrease* in amplifier gain also may cause instability. Such systems are still covered by the polar plot criterion of stability, and are commonly called "Nyquist stable," or "conditionally stable."

$-\mu\beta$ exhibits gain and phase margins of the order of 10 db and 50 degrees respectively, the transient response of the modifying factor to a step function will be well-damped and generally not overshoot more than about 25%. If the gain margin is sufficient, the phase margin usually will be the dominant factor in determining the size of the initial overshoot. The required phase margin for critical damping depends upon the exact shape of $\mu\beta(j\omega)$, but in general is about 60 degrees. The gain margin needed in a particular design will depend upon the expected variability of the loop transmission. Radar tracking loops should usually have gain margins of the order of 15

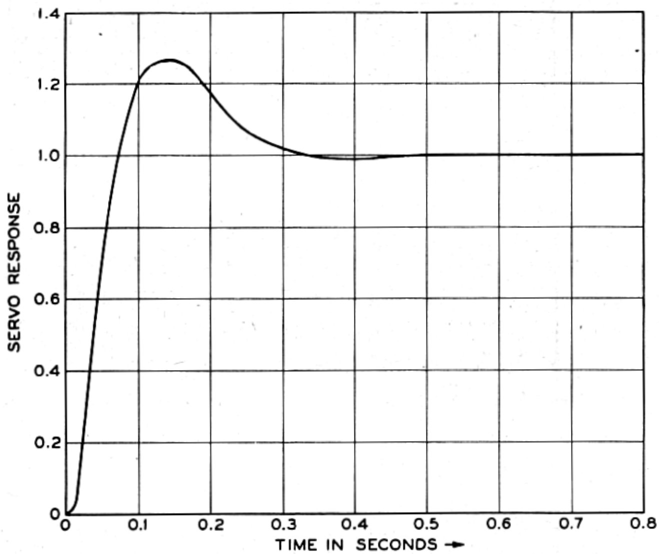

Fig. 9—Transient response of illustrative servo system.

db or more because of the large number of factors which may cause the loop gain to vary.

While the polar diagram gives a clear picture of stability considerations, it is usually more convenient for design purposes to plot the gain and phase of $-\mu\beta$ as separate curves on a logarithmic frequency scale, for positive values of ω. This is illustrated in Fig. 10, for the sample servo system. Under two commonly met conditions, the requirement for single loop[11] stability on this type of plot is simply that the absolute value of phase shift be less than 180 degrees at zero db gain ($|\mu\beta| = 1$). The conditions are that the connective polarity be such as to make $-\mu\beta$ positive when the

[11] Again, multiple loop systems may be included if all subsidiary loops are individually stable.

network phase shifts vanish, and that the gain curve cross zero db at only one frequency.[12]

An advantage of this logarithmic diagram is that commonly encountered forms of $|\mu\beta|$ vary as $\omega^{\pm n}$ for intermediate or asymptotic frequency regions, and thus plot as corresponding straight line segments. From (7.4) it may be seen that the illustrative form of $|\mu\beta|$ behaves, in turn, as $200\omega^{-1}$, $200\omega^{-2}$, $20\omega^{-1}$, and $1.6 \times 10^8\omega^{-4}$, as ω is increased. These asymptotic lines are drawn in lightly in Fig. 10, the actual gain describing smooth transi-

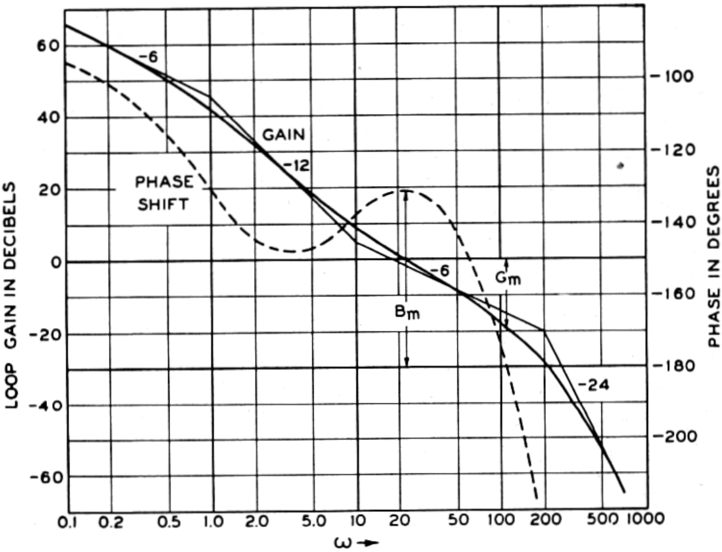

Fig. 10—Loop characteristic of illustrative servo system.

tions between adjacent asymptotes. Since the logarithmic slope of $\omega^{\pm k}$ is $\pm 6k$ db/octave,[13] the successive asymptotic slopes in Fig. 10 are -6, -12, -6, and -24 db/octave. The junctures of adjacent asymptotes occur at values of ω of 1, 10, and 200. These juncture frequencies are called "corner" frequencies, and may be seen from (7.2) to coincide with the real constants or "roots" added to $j\omega$ in each of the three factors in parentheses. The corner associated with the factor $200/j\omega$ is of course at $\omega = 0$. The corner of the last, or cubed factor is a multiple one, joining two asymptotes differing in slope by 18 db/octave. From a knowledge of such corner

[12] This discussion assumes that $\mu\beta$ has a low-pass configuration; that is, only the high frequency cut-off is considered. If $\mu\beta$ has a ow-frequency cut-off also, then a corresponding requirement identical with the above must be added.

[13] An octave is taken to be a 2:1 span in frequency, and 6 db is of course very closely equivalent to a 2:1 increase in $|\mu\beta|$.

frequencies, and the fact that the actual gain curve lies 3 db from an isolated simple asymptotic corner, the gain curve can usually be drawn in without further computation.[14] The phase curve also is easily constructed by adding up the elementary phase curves associated with the various corners. As may be seen from (7.3), these component phase curves all will have the same shape on a logarithmic frequency plot, merely being shifted along the frequency scale. The phase contributed by each simple corner will be $\pm$ 45 degrees at the corner frequency, the sign depending upon whether the associated root appears in the numerator or in the denominator.

It is an extremely important fact that the very requirement of stability imposes an unambiguous interrelationship between the gain and the phase shift of most types of transfer characteristic! By the general mathematical methods leading to the previously discussed stability criteria, Bode[15] has shown that this is true for the broad class of network structures commonly used in feedback loop design. That is, if either the transfer gain or phase shift is specified *at all frequencies*, the attendant phase or gain can be computed without further information. This class of networks is called *minimum phase*. Any stable structure composed of lumped circuit elements will have a transfer characteristic of the minimum phase type, provided it does not include an all-pass section.[16] All-pass characteristics are seldom used in the design of feedback loops, since their inclusion in the loop always reduces the stability margins achievable with a given high-frequency cut-off. Thus the unique interrelationship between phase and gain may be assumed for the loop characteristic $-\mu\beta$ in single-loop feedback systems. The nature of this relationship is discussed in detail by Bode. Briefly, the phase shift at any frequency ω_c is proportional to a weighted average of the gain slope in db/octave, over the entire logarithmic frequency scale. The weighting factor sharply emphasizes gain slopes in the immediate vicinity of ω_c, while the contributions of gain slopes at remote frequencies are reduced in proportion to the logarithmic frequency span from the particular frequency ω_c.[17] For transfer characteristics of the form $\omega^{\pm k}$, having a constant gain slope of $\pm 6k$ db/octave,[18] the associated phase shift is also constant and equal to $\pm 90k$ degrees. For transfer functions which behave approximately as $\omega^{\pm k}$ over a finite frequency span, the phase shift

[14] The corner frequency concept is less useful if the roots are complex. However a great many servo systems are so constructed that $\mu\beta$ has only real roots.

[15] Loc. cit. Also see "Relations between attenuation and phase in feedback amplifier design," by H. W. Bode, *B. S. T. J.*, July 1940, p. 421.

[16] An all-pass section is one which has constant gain but varying phase shift versus frequency, and is usually composed of a lattice, bridged T, or other bridge type circuit.

[17] About 60% of the area under this weighting function lies between $\omega = 0.5\ \omega_c$ and $\omega = 2\ \omega_c$, 80% between 0.25 ω_c and 4 ω_c.

[18] That is, for transfer characteristics whose *absolute magnitude* is given by $\omega^{\pm k} \cdots$.

of $\pm 90k$ degrees is approached more and more closely as the length of span is increased.

This may be observed qualitatively from the transfer characteristic of Fig. 10. For $\omega \ll 1$, the gain slope is -6 db/octave, and the phase shift approaches -90 degrees. For $1 < \omega < 10$, the average gain slope is about -10 db/octave, and the phase shift near $\omega = 3$ is -148 degrees (instead of $-90 \times 10/6 = -150$ degrees). As ω increases toward 200, the phase shift increases rapidly due to the asymptotic slope of -24 db/octave, finally approaching -360 degrees ($-90 \times 24/6$) for $\omega \gg 200$.

Foreknowledge of the inevitable gain-phase relationship is of great value to the servo designer, in making clear the comparatively small class of realizable gain-phase combinations and thus averting attempts at non-physical designs. For example the design use of too-rapidly falling loop gain characteristics in the region of the high-frequency gain cross-over (that is, near zero db loop gain) is not permissible because of the large negative phase shifts which must accompany the steep gain slopes. Another way of stating the advantage of an early realization of the gain-phase laws is to say that the designer is assured in advance that any paired gain and phase characteristics which he chooses within the basic restrictions will be achievable with stable physical networks.[19]

3.2 *Dynamic Error*

A servo system is usually designed to transmit some class of input functions with a required degree of fidelity. This class of functions may reduce substantially to one specific input signal whose time variation or whose frequency spectrum is known, or it may include a great variety of signals which have certain properties in common. In the latter case it is conceivable that definite limits may be placed upon the allowable amplitude ranges of the input signal and its various time derivatives, or certain limiting frequency spectrum characteristics may be specified for the input function.

Servomechanisms are subject to several types of transmission error. The systematic error, or dynamic error, is predictable from knowledge of the *noise-free* input signal and of the transfer frequency characteristic of the servo system. For simplicity, the discussion of error will be limited to the case where the output signal is desired to be a replica of the input, and where $\beta = -1$. Thus the loop transmission $\mu\beta$ becomes simply $-\mu$. The input-output relationship as given by (6) is therefore

$$F_2 = \frac{\mu}{1 + \mu} F_1, \tag{9}$$

[19] With some necessary reservations as to practicable dissipation constants and parasitic circuit constants.

where F_1 and F_2 are again typical sinusoidal components of the input and output respectively. Thus the corresponding *sinusoidal error component* may be written as

$$\Delta = F_1 - F_2 = \frac{F_1}{1 + \mu}. \tag{9.1}$$

The methods which may be used to determine the actual dynamic error $\Delta(t)$ from (9.1) depend both upon the nature of $f_1(t)$ and the type of information available about $f_1(t)$. If the input signal is a known periodic function, $\Delta(t)$ may be found by applying (9.1) for each sinusoidal component of the input and summing the resulting terms. If the input is non-periodic in character, then the error may be calculated from the Fourier integral expression

$$\Delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{F_1(\omega)}{1 + \mu} e^{j\omega t} \, d\omega, \tag{10}$$

where $F_1(\omega)$ represents the continuous frequency spectrum of $f_1(t)$, as obtained from

$$F_1(\omega) = \int_{-\infty}^{\infty} f_1(t) e^{-j\omega t} \, dt. \tag{10.1}$$

The problems of calculating $F_1(\omega)$ from $f_1(t)$ and $\Delta(t)$ from $F_1(\omega)$ often may be avoided by consulting well-known tabular lists of paired time and frequency functions.[20]

Equation (9.1) may be used as a broad guide in selecting the type of μ characteristic best suited to a particular input signal. It has been mentioned previously that because of input noise and parasitic circuit elements, the servo transfer bandwidth usually should be kept as narrow as possible, consistent with dynamic error requirements. The transfer characteristic $\mu/(1 + \mu)$ will be closely equal to unity while $|\mu| \gg 1$, will rise slightly[21] in the region where $|\mu| \backsimeq 1$, and fall off as μ when μ is small compared with unity. The "cross-over" frequency, for which $|\mu| = 1$, may be taken as a rough measure of the transfer bandwidth. Thus, the requirement of minimizing the bandwidth may be restated as that of minimizing the cross-over frequency, while holding the dynamic error within specified limits. Reasoning in a general way, this requirement may be met by designing μ so that the amplitudes of the sinusoidal error components, as given by

[20] An excellent list is given by G. A. Campbell and R. M. Foster in a Bell System monograph "Fourier Integrals for Practical Application," September, 1931. A table of Laplace Transforms, which also may be used, is given by M. F. Gardner and J. L. Barnes in "Transients in Linear Systems," John Wiley and Sons Inc., 1942.

[21] Assuming a phase margin of the order of 60 degrees.

(9.1), are roughly constant with frequency over the servo band. *This demands that μ have somewhat the same frequency distribution as the input signal spectrum* (for $|\mu| \gg 1$). Because of stability requirements and complexity of the necessary apparatus, this rule can usually be followed over only a part of the servo frequency band, especially when the input signal spectrum falls off very rapidly with increasing frequency. However, even a rough adherence to this desired relation is usually of real worth in reducing the noise errors of the servo. An illustration of this will be given in a later section.

3.21 *Approximate Calculation of Dynamic Error*

Frequently the servo requirement is to transmit, with great accuracy, a type of signal whose frequency spectrum falls off very rapidly with increasing frequency. As may be seen from (9.1) this demands very large values of loop transmission μ at the lower frequencies where the input signal energy is concentrated, but permits a rapidly dropping loop transmission versus frequency commensurate with the falling amplitude spectrum of the input signal. Such a rapid reduction in loop gain is practicable while $|\mu| \gg 1$. However, stability considerations force a more gradual gain reduction as the region of gain cross-over is approached. As a result, contributions to the servo error from this frequency region may be neglected compared with those from the lower frequencies. This suggests a series expansion of (9.1) in the form,

$$\Delta = [a_0 + a_1(j\omega) + a_2(j\omega)^2 + a_3(j\omega)^3 + \cdots] F_1 , \tag{11}$$

where a_0, a_1, etc. are real constants.

Because of the assumed rapid drop in component amplitude F_1 with increasing frequency it is often unnecessary to take account of more than a few terms of the expansion.[22]

It is easy to show that (11) may be rewritten on a time basis to give the total dynamic error as

$$\Delta(t) = a_0 f_1(t) + a_1 \dot{f}_1(t) + a_2 \ddot{f}_1(t) + \cdots , \tag{12}$$

where $(\dot{\ }) = d(\ \ )/dt$. Thus the coefficient a_0 gives the error component proportional to input displacement. Similarly, a_1 and a_2 are the coefficients of the error components due to input velocity and input acceleration, respectively. For a great many motor-drive servo systems the loop transmission μ approaches infinity as $1/j\omega$ when ω approaches zero. This en-

[22] The series may be said to converge rapidly in a practical sense, for the following reason: For small values of ω the higher order terms are negligible. For values of ω sufficiently large that the high order terms may no longer be neglected the coefficient F_1 has become so small as to make the contribution of the entire series negligible.

sures that a_0 and thus the displacement error will be zero, leaving principally the velocity and acceleration errors to be considered.

The coefficients a_0, a_1, a_2, etc. may be calculated easily for any particular case. For illustration, the three common forms of μ characteristic shown in Fig. 11 will be examined. (As previously discussed, the designated forms of μ need hold only for $|\mu| \gg 1$.)

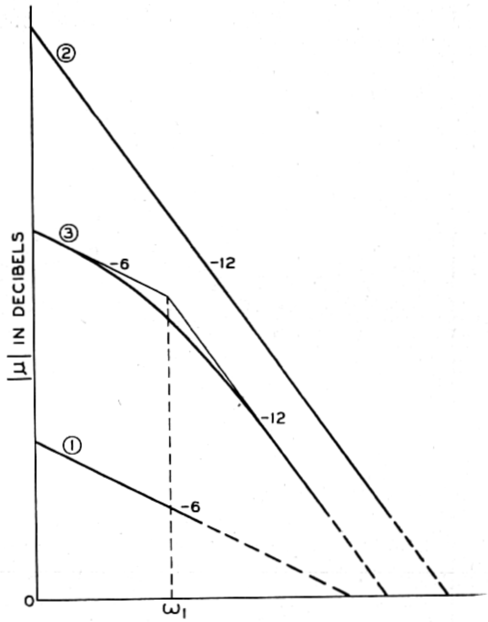

Fig. 11—Elementary μ forms.

Type 1. -6 *db/octave,* $\mu = \omega_0/j\omega$

The error expansion becomes,

$$\Delta(t) = \underline{\frac{1}{\omega_0}\dot{f}_1(t)} - \frac{1}{\omega_0^2}\ddot{f}_1(t) + \cdots . \quad (12.1)$$

For the combination of high accuracy and rapidly converging input spectrum, the first term is the only one of importance. Thus this type of system has essentially a *velocity* error.

Type 2. -12 *db/octave,* $\mu = (\omega_0/j\omega)^2$

Here the error is

$$\Delta(t) = \underline{\frac{1}{\omega_0^2}\ddot{f}_1(t)} - \frac{1}{\omega_0^4}\ddddot{f}_1(t) + \cdots . \quad (12.2)$$

Again for the rapidly converging case, this system will have principally an *acceleration* error.

Type 3. -6, -12 *db/octave*, $\mu = \omega_0\omega_1/j\omega(j\omega + \omega_1)$

This is perhaps the most commonly encountered characteristic in simple servos. The corresponding error expansion is

$$\Delta(t) = \underline{\frac{1}{\omega_0}\dot{f}_1(t)} + \underline{\frac{1}{\omega_0\,\omega_1}\ddot{f}_1(t)} - \frac{2}{\omega_0^2\,\omega_1}\dddot{f}_1(t) - \cdots, \qquad (\omega_0 \gg \omega_1). \quad (12.3)$$

and the principal error for this type system thus is a combination of *velocity* and *acceleration* components. Either the velocity or the acceleration error component may be predominant, depending upon the various parameters.

3.3 *Noise Errors*

The typical sinusoidal component of servo error due to noise (unwanted signals or irregularities) in the input signal may be written as*

$$\Delta_n = \frac{\mu}{1 + \mu} N, \quad (13)$$

where N represents the corresponding sinusoidal component of the input noise. If the noise signal $n(t)$ is known, the total noise error $\Delta_n(t)$ may be calculated from (13) in the ways described for the dynamic error. However, the noise input is seldom known in this sense, although certain outstanding components sometimes may be estimated and their effects evaluated. On the other hand the average disturbance due to random input noise, of the kind described as "thermal noise" in electrical circuits, may easily be calculated. This type of noise has constant amplitude versus frequency, and the total power in the output noise error may be found from

$$P_n = K \int_0^\infty \left| \frac{\mu}{1 + \mu} \right|^2 d\omega, \quad (14)$$

where K is a constant dependent upon the input noise power.

Input noise also causes overloading of the power amplifier and overheating of the motor. These effects are aggravated by the falling transfer characteristic versus frequency of the motor, as seen from the following discussion. The servo transfer characteristic is maintained approximately at unity out to the cross-over frequency. However the transfer ratio of the motor, equation (3.1), will be falling at least at 6 *db*/octave, usually at 12 *db*/octave, at frequencies below this point.[23] Thus the transfer

* Again assuming $\beta = -1$.

[23] Assuming that the mechanical load impedance is a series combination of resistance and inertia.

characteristic (loop closed) from the servo input up to the motor and power amplifier must rise correspondingly with frequency, out to the cross-over point. Again assuming input noise of the uniform amplitude versus frequency type, the total noise power at the motor input is therefore,

$$P_{nm} = K_1 \int_0^\infty \left| \frac{\mu}{1+\mu} \right|^2 (\omega^2 + \omega_m^2)\omega^2 \, d\omega. \tag{15}$$

Again, ω_m is the reciprocal time-constant of the motor and K_1 is a proportionality constant. If ω_m is less than about half the cross-over frequency, then the noise power at the motor input increases as the fifth power of the bandwidth of the servo transfer characteristic.[24] Thus, if the input signal/noise ratio is small, this effect may be an important design consideration.

Still other servo errors may result from local extraneous signals or from coulomb and static frictional effects. These error sources are in a somewhat different class from those discussed previously, in that they are more nearly under the designer's control. That is, such extraneous signals and friction may be kept small by proper design and the residual friction effects further reduced by the use of local feedback. In the absence of local feedback, the servo error resulting from frictional or other torque disturbances at the output shaft readily is found to be

$$\Delta_T = \frac{T}{S(j\omega)} \cdot \frac{1}{1+\mu}. \tag{16}$$

$S(j\omega)$ is the actual stiffness (loop opened) of the output mesh, and T is the disturbing torque. T conceivably may represent static or coulomb friction, load-torque irregularities due to fluctuating running-friction, or wind torque. Again assuming the mechanical impedance to be resistance and inertia in series, the mechanical stiffness is, from (2.2), $S(j\omega) = j\omega(R + j\omega J)$. Thus the error is

$$\Delta_T = \frac{T}{j\omega(R + j\omega J)} \cdot \frac{1}{1+\mu}, \tag{16.1}$$

and the apparent output stiffness (loop closed) is

$$S' = j\omega(R + j\omega J)\,(1 + \mu). \tag{16.2}$$

If T is taken as the static load torque, the resulting static error is found by setting $\omega = 0$ in (16.1). Assuming that μ behaves as $\omega_0/j\omega$ when ω approaches zero, the static error is

$$\Delta_T = \frac{T}{\omega_0 R}, \tag{16.3}$$

[24] This assumes a constant functional form for the transfer characteristic. However, the statement holds approximately, even with considerable variation in this form.

and the apparent low-frequency stiffness is $\omega_0 R$, being the ratio of the mechanical resistance to the velocity error coefficient. It may be noted that the static error will vanish if the loop transmission approaches infinity more rapidly than $1/\omega$ as ω approaches zero.

3.4 *Comparison of μ Characteristics for a Particular Input Signal*

In order to illustrate the advantages of shaping the loop characteristic for a particular input signal, a brief discussion will be given of the design of an automatic radar loop to track an airplane in azimuth over a constant linear-velocity course. The servo configuration is that given by Fig. 5b, θ_1 being the azimuth angle of the target and θ_2 the corresponding antenna angle. The lobing radar antenna has been assumed to take the place of the

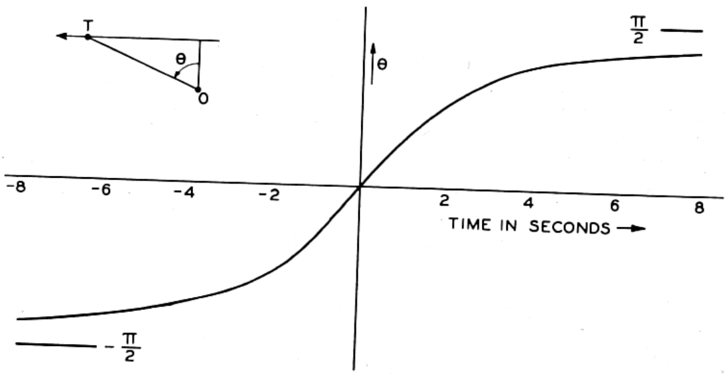

Fig. 12—Azimuth angle for constant linear velocity course.

synchro pair. Thus $\beta = -1$, and an error signal proportional to $\theta_1 - \theta_2$ is developed.[25]

Assuming a constant linear-velocity course having a maximum azimuth rate of 30 degrees/sec, the target azimuth angle is given by [26]

$$\theta_1(t) = \tan^{-1} .524t, \tag{17}$$

which is plotted in Fig. 12.

This course will develop a maximum azimuth acceleration $\ddot{\theta}_1$ of ± 10.3 degrees/sec^2 and a maximum $\dddot{\theta}_1$ of -16.4 degrees/sec^3. The continuous frequency spectrum of $\theta_1(t)$ may be found from (10.1) to be

$$F_1(\omega) = \pi \frac{e^{-1.9|\omega|}}{j\omega}. \tag{18}$$

[25] Assuming a low elevation course.
[26] The azimuth angle has been so taken that zero azimuth is obtained at the point of nearest approach.

A logarithmic plot of $| F_1(\omega) |$ is shown in Fig. 13.* It may be seen that the input signal spectrum falls at 6 *db*/octave for $\omega \ll 0.5$, at 12 *db*/octave for $\omega = 0.524$, and 30 *db*/octave at $\omega = 2.1$.

Assuming that the permissible dynamic error is 0.3 degree, a comparison will be made between the type 1 and type 3 loop characteristics of the previous section. For the type 1 system, which will have essentially a pure velocity error, (12.1) shows the required value of ω_0 to be 30/0.3 or 100.

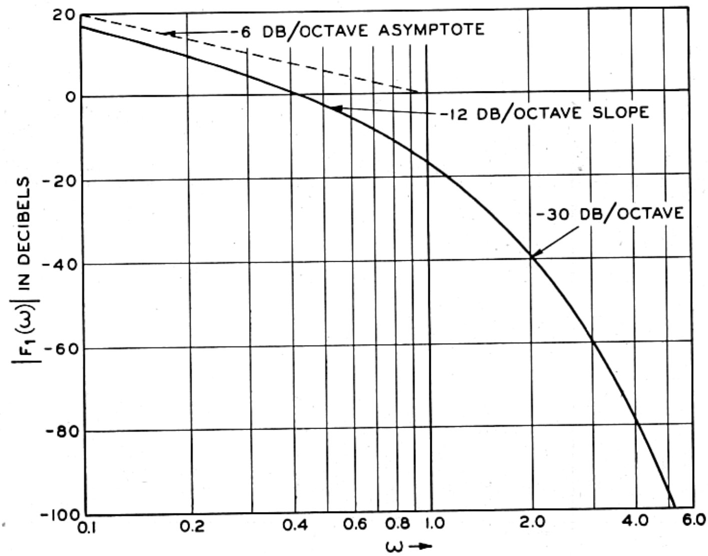

Fig. 13—Target frequency spectrum for constant velocity course.

Thus $\mu = 100/j\omega$. Figure 14 shows a logarithmic plot of the corresponding $| \mu |$. This characteristic departs rapidly from the shape of input signal spectrum given by Fig. 13, as ω is increased abcve 0.1.

The type 3 characteristic permits a considerably better match. Choosing a compromise value for ω_1 of 0.1, (12.3) may be used to calculate the necessary value of ω_0 as 415. Thus the loop transmission becomes $\mu = 41.5/j\omega(j\omega + 0.1)$. Figure 14 shows a plot of the corresponding $| \mu |$, modified near the gain cross-over to satisfy the stability requirements. This curve is a considerably better average match for the target frequency-spectrum up to $\omega = 1$. The resulting type 3 system has a predominant acceleration error as judged from the maximum velocity and acceleration errors of .072 degree and 0.25 degree respectively.

The total dynamic error curves for the constant-velocity course are given

* $| F_1(\omega) | = \pi$ has been taken as the zero *db* level.

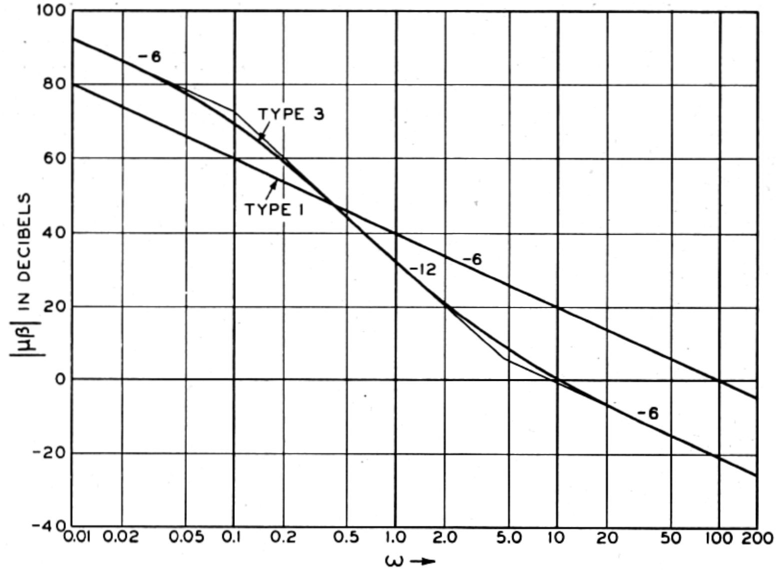

Fig. 14—Tracking loop characteristics.

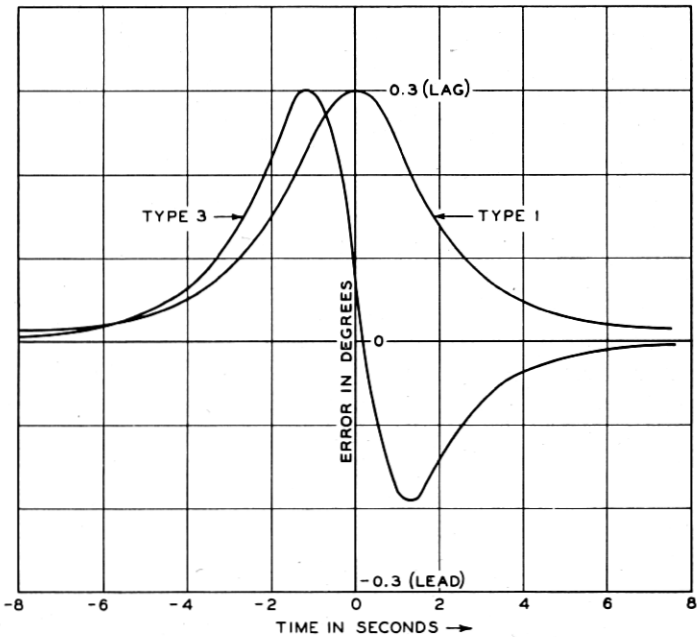

Fig. 15—Tracking errors for constant velocity course.

in Fig. 15. The velocity error of the type 1 system is always a lagging error and is maximum at the point of nearest approach. The type 3 composite of velocity and acceleration errors is lagging over about the first half of the course and leading for the second half, having lead and lag maxima at points closely grouped about the point of nearest approach.

Although the two loop characteristics develop the same maximum dynamic error on the specified target course, their transient responses to an input step differ widely, as may be seen from Fig. 16. The rise time for the type 1 loop is about .03 second compared with an initial rise in 0.17 second

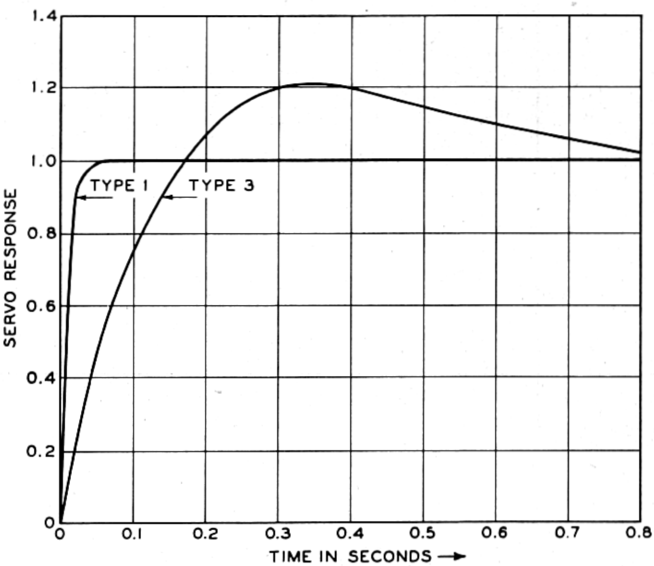

Fig. 16—Transient response of tracking servos.

for the type 3 system. Also, because of the overshoot the type 3 system requires about 0.7 second to settle within 5% of the equilibrium value.

For a final comparison of the two systems the corresponding transfer characteristics, $\mu/(1 + \mu)$, are plotted in Fig. 17 on arithmetic amplitude and frequency scales. It may be seen that the type 1 system is vulnerable to noise and interfering signals over a far wider frequency range than the type 3. Again assuming uniform input noise versus frequency, (14) may be used to show that the ratio of output noise power for the two systems is about 7.5:1.

Thus the luxury of crisp transient response as obtained with the type 1 system may demand a heavy penalty in terms of output fluctuations due to noise and other unwanted signal variations. This is a clear illustration of the necessity for designing the servo loop to match the type of input signal to be transmitted, particularly for radar tracking systems where the "unwanted variations" are ever present.

3.5 *Use of Local Feedback*

There are many examples of the use of local or subsidiary feedback in servo systems. The more common of these include feedback around vacuum tube power amplifiers to obtain improved linearity and impedance properties, and over-all feedback around amplifier and motor-drive systems to

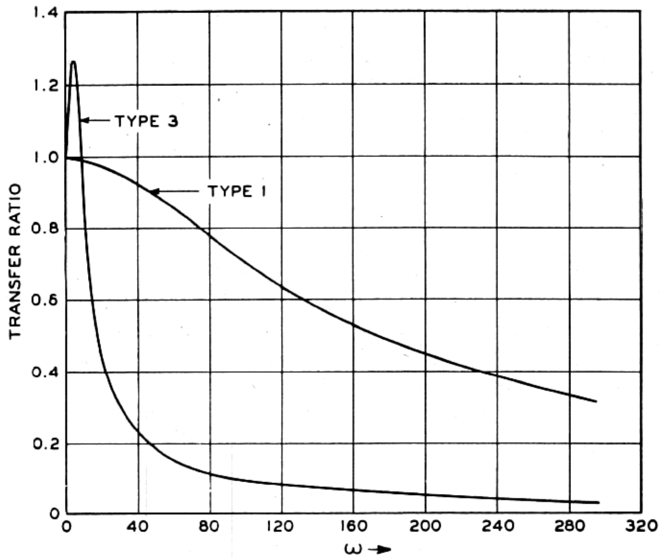

Fig. 17—Frequency response of tracking servos.

suppress frictional effects, increase output stiffness, and modify the inherent frequency characteristics of the basic components.[27] The tendency toward "β circuit dependency" as previously discussed also produces greater constancy of the stage transfer characteristics with time, temperature, etc.

Perhaps the simplest and most useful kind of local feedback is negative tachometer (velocity) feedback around motor-drive systems. This type of feedback widens the transfer frequency band of the drive system by reducing its time-constant, and increases the linear speed range of the motor. This may be illustrated by referring back to Fig. 7, which shows a typical tachometer loop. Assuming the transfer ratio of the amplifier to be a con-

[27] In a slightly different class are the servo systems used to provide automatic frequency and gain control in radio systems.

stant A, the transfer ratio of the motor and amplifier without feedback is, from (3.1),

$$\mu_T = \frac{\theta}{E} \text{ (loop open)} = \frac{\mu_0}{J} \frac{1}{j\omega(j\omega + \omega_m)}, \tag{19}$$

where the constant $\mu_0 = A\mu_t$. (To avoid confusion with primary loop quantities, the tachometer loop will be represented by the symbols μ_T and β_T, rather than μ and β.) The quantity ω_m was defined as the ratio $(R_m + R_m')/J$ (see Fig. 2b), and is the reciprocal of the motor time constant. Replacing $(R_m + R_m')$ by R for convenience, (19) may be rewritten as

$$\mu_T = \frac{\theta}{E} \text{ (loop open)} = \frac{\mu_0}{j\omega(R + j\omega J)}. \tag{19.1}$$

The transfer ratio of the tachometer is

$$\beta_T = \frac{E_\beta}{\theta} = -j\omega R_t,$$

and thus the loop transmission characteristic is

$$\mu_T \beta_T = -\frac{\mu_0 R_t}{R + j\omega J}. \tag{20}$$

For values of ω small compared with ω_m this loop transmission is constant and closely given by $\mu_T\beta_T(0) = -\mu_0 R_t/R$. When $\omega \gg \omega_m$, $\mu_T\beta_T$ approaches the form $-\mu_0 R_t/j\omega J$, and thus falls off at 6 *db*/octave. Consequently the maximum phase shift of the factor $-\mu_T\beta_T$ is -90 degrees, and no stability problem arises for the local tachometer loop.[28]

From (19.1) and (20), the over-all transfer ratio with feedback is

$$\begin{aligned} \frac{\theta}{E} \text{ (loop closed)} &= \frac{\mu_T}{1 - \mu_T \beta_T}, \\ &= \frac{\mu_0}{j\omega(R + \mu_0 R_t + j\omega J)}. \end{aligned} \tag{21}$$

Comparing (21) with (19.1), it may be seen that the sole effect of the tachometer feedback upon the over-all transfer ratio has been to add an apparent "ohmic" friction or mechanical resistance $\mu_0 R_t$ to the original value R. (It will be shown that this increase in apparent mechanical resistance also is effective in increasing the mechanical output impedance, although no power is dissipated in the added component $\mu_0 R_t$.)

[28] Actually, the effects of parasitic elements always modify this situation somewhat, especially if unusually high loop transmission is sought. However tachometer loops often require little or no stabilizing equalization.

Equation (21) also may be written as

$$\frac{\theta}{E} \text{ (loop closed)} = \frac{\mu_0}{J} \frac{1}{j\omega(j\omega + \omega_m')}, \tag{21.1}$$

where $\omega_m' = (R + \mu_0 R_t)/J$ is the new corner frequency.

The change in over-all transfer ratio due to the tachometer feedback is shown in Fig. 18. The solid line diagrams A and B are the transfer gains without feedback and with feedback, respectively.[29] At low frequencies such that $\omega \ll \omega_m$, the feedback reduces the transfer ratio by the factor ω_m'/ω_m, the ratio of the two corner frequencies.[30] In order to restore this low-frequency loss in transmission, it is necessary to provide an added

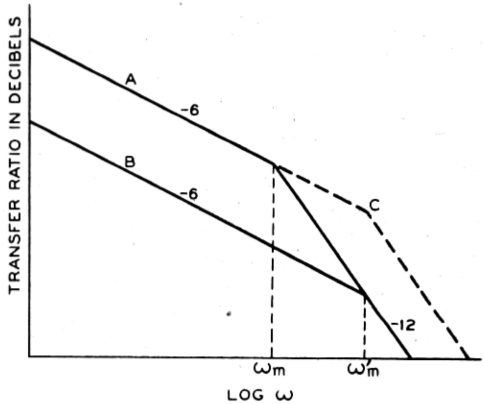

Fig. 18—Effect of tachometer feedback on motor characteristic.

amplification ω_m'/ω_m. If this is accomplished by increasing μ_0 and decreasing R_t so that the product $\mu_0 R_t$ remains constant,[31] the resulting transfer ratio will be that shown by the dotted lines C in Fig. 18. Comparing A and C, it may be seen that the net result of applying tachometer feedback and increasing the amplifier gain is to widen the transfer bandwidth by the factor ω_m'/ω_m. The required increase in amplification is the cost of widening the transfer bandwidth either by tachometer feedback or by non-feedback means, such as the use of a "forward-acting" equalizer in the amplifier. (However, such forward acting equalization fails to provide the increased over-all linearity and mechanical impedance obtained by the feedback method.) At frequencies sufficiently high that $\omega \gg \omega_m'$, the change in transfer ratio due to the feedback disappears, the mechanical inertia becoming the controlling element.

[29] The straight line asymptotes have been drawn instead of the actual gain curves.

[30] This is also the factor by which the feedback reduces the output speed obtained for a steady input voltage, neglecting circuit non-linearites and coulomb friction.

[31] This ensures a fixed loop transmission, and thus an unchanging value for ω_m'.

For ω small compared with ω_m', (21) becomes

$$\frac{\theta}{E}\,(\text{loop closed}) \simeq \frac{\mu_0}{j\omega(R + \mu_0 R_t)}, \qquad (\omega \ll \omega_m').$$

If the tachometer feedback is substantial ($\omega_m' \gg \omega_m$), this may be further approximated as

$$\frac{\theta}{E}\,(\text{loop closed}) \simeq \frac{1}{j\omega R_t}, \qquad \begin{pmatrix} \omega \ll \omega_m' \\ \omega_m' \gg \omega_m \end{pmatrix}. \tag{21.2}$$

and the corner frequency becomes

$$\omega_m' \simeq \frac{\mu_0 R_t}{J}, \qquad (\omega_m' \gg \omega_m).$$

Thus for reasonably high feedback, the over-all transfer ratio (21.2) depends only upon the tachometer characteristic, being substantially independent of changes in the original mechanical resistance R or the amplifier-motor factor μ_0. The corner frequency ω_m' is similarly independent of changes in R, although still a direct function of μ_0. Thus the principal non-linearity of two-phase induction motors, namely variation in electrical damping with speed, is effectively suppressed by this type of local feedback, and systems employing such motors up to 80% of their synchronous speed may be designed on a linear basis.

The increase in mechanical impedance due to the feedback may be shown by assuming a torque disturbance T applied at the output shaft. Without feedback, the resulting speed disturbance is

$$\dot{\theta}\,(\text{loop open}) = \frac{T}{Z_m} = \frac{T}{R + j\omega J}.$$

With feedback, the corresponding shaft speed disturbance becomes

$$\dot{\theta}\,(\text{loop closed}) = \frac{T}{Z_m} \cdot \frac{1}{1 - \mu_T \beta_T},$$

$$= \frac{T}{R + \mu_0 R_t + j\omega J}.$$

Thus the apparent mechanical resistance, and therefore the protection against frictional torques, has been multiplied by a factor $(1 + \mu_0 R_t/R) = \omega_m'/\omega_m$. If the motor-drive system with tachometer feedback is employed in a simple follow-up system of the type of Fig. 5, equation (16.3) shows that the resulting low-frequency output-shaft stiffness will be $\omega_0(R + \mu_0 R_t)$ or $(\omega_m'/\omega_m)\omega_0 R$.[32] Therefore the output stiffness has

[32] The low-frequency loop transmission of the follow-up loop is again taken to be $\omega_0/j\omega$.

been increased by the factor ω_m'/ω_m over that obtained without the use of local feedback, assuming identical follow-up loop characteristics ($\mu\beta$) for the two cases. The ratio ω_m'/ω_m thus directly measures the feedback eduction of static and low-speed errors of the follow-up system due to torque disturbances. In practice the resulting increase in static accuracy may be of the order of 10 to 100 times.

3.6 *Error Reduction by Non-Feedback Means*

In situations where the noise associated with the input signal is small, it may be desirable to reduce the dynamic errors obtained with a given servo system by the use of forward-acting equalization external to the loop.

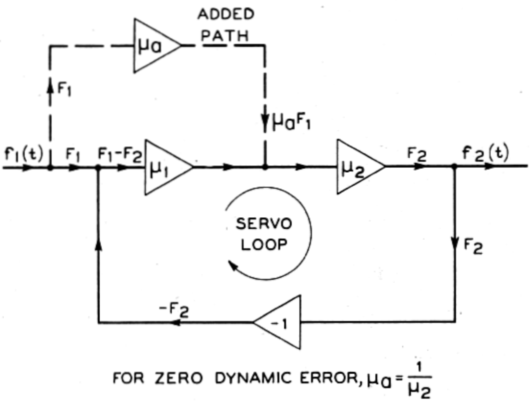

Fig. 19—Forward-acting error compensation.

That is, the dynamic error characteristic may be computed, and the servo input or output modified by supplementary networks in such a fashion as to reduce the over-all error.

An illustrative arrangement, which is suitable when the input member is accessible,[33] is shown in Fig. 19. For convenience the servo is taken to be a simple follow-up system having $\beta = -1$. The μ circuit is shown divided into two parts, μ_1 and μ_2. Typically, μ_1 may be the transfer stiffness of a synchro pair (Fig. 3b), and μ_2 the transfer characteristic of a motor-drive system. The normal dynamic error component for such a loop, omitting the dotted line, has been shown to be $F_1/(1 + \mu)$. If an additional signal $\mu_a F_1$ is obtained from the input member and injected into the system as shown by the dotted line, then

$$F_2 = \frac{\mu}{1+\mu} F_1 + \frac{\mu_a \mu_2}{1+\mu} F_1,$$

$$= \frac{\mu + \mu_a \mu_2}{1+\mu} F_1.$$

[33] This is not the case for a radar tracking loop, for instance.

Thus the over-all error becomes

$$F_1 - F_2 = \left(1 - \frac{\mu + \mu_a \mu_2}{1 + \mu}\right) F_1,$$

or

$$F_1 - F_2 = \frac{1 - \mu_a \mu_2}{1 + \mu} F_1. \tag{22}$$

If the added transmission path is so designed that

$$\mu_a = \frac{1}{\mu_2}, \tag{23}$$

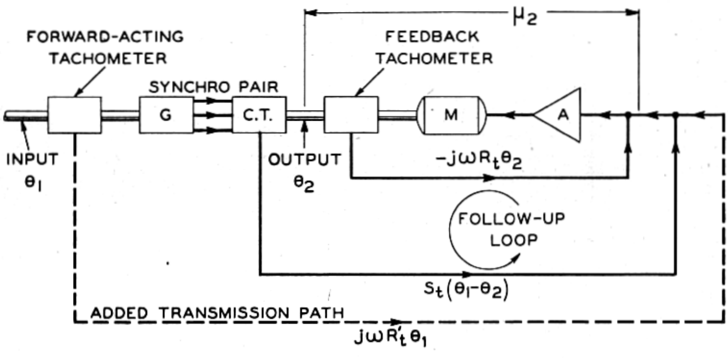

Fig. 20—Forward-acting tachometer system.

then $F_1 = F_2$, and the dynamic error vanishes. Thus the desired form of the added transmission depends only upon the μ_2 portion of the loop characteristic. It will not be possible to satisfy the condition given by (23) exactly, especially at the higher frequencies where noise enhancement and parasitic effects will become increasingly important. However, it is often possible to obtain the proper form for μ_a over the range of frequencies responsible for the bulk of the dynamic error. If μ_a has the proper frequency characteristic but is too large by 10%, for instance, it may be seen from (22) that there still remains a 10/1 increase in dynamic accuracy.

The foregoing method is especially applicable when μ_2 represents the transfer characteristic of a motor-drive system employing tachometer feedback, as shown in Fig. 20. Here the basic input-output comparison is obtained by means of the synchro pair, while a tachometer coupled to the input shaft provides the error-reducing signal. Thus the transmission μ_a is equal to $j\omega R_t'$, where R_t' is the tachometer transfer resistance. The expression for

μ_2 is given approximately by (21.2) as $1/j\omega R_t$. Thus, by (23), $R_t' \simeq R_t$ for substantial cancellation of the dynamic error (at frequencies small compared with ω_m'). That is, the output voltages of the two tachometers must closely annul each other when the input and output shafts are travelling at the same speed. Since the tachometers may be closely alike and excited from the same supply line, it is comparatively easy to keep their transfer ratios closely matched. In practice an error reduction of 20/1 is readily maintained by this method.

The error compensation scheme described above does not change the loop characteristic $\mu\beta$ of the basic servo loop, and thus does not create new stability problems. Its use to obtain high servo accuracy is desirable when the input noise is small and when a high loop gain is difficult to obtain because of parasitic elements or equipment complexities.